THE UNIVERSITY OF ALBERTA

A MODE-SPLIT MODEL FOR COMPULSORY TRIPS

BASED ON ECONOMIC THEORY

by

© ALOIS PETER KOLINSKY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF CIVIL ENGINEERING

EDMONTON, ALBERTA

Spring, 1971

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled A MODE-SPLIT MODEL FOR COMPULSORY TRIPS BASED ON ECONOMIC THEORY submitted by ALOIS PETER KOLINSKY in partial fulfilment of the requirements for the degree of Master of Science.

ABSTRACT

This thesis investigates the relationship between the mode-split, (i.e. the ratio of demand for one mode of passenger transport to the demand for all modes), for trips of necessity and its principal influential factors. The investigation is warranted because compulsory trips in the form of work trips comprise a large proportion of the total number of intracity passenger trips, especially during peak hours. A current passenger travel demand model, which is founded on the economic theory of demand, is introduced and criticized with respect to compulsory trips. A binary choice model is then introduced which gives an indirect measure of mode-split. What the binary choice model does give is the probability that a trip will be conducted by a certain mode where the choice is between two modes.

The method for calibrating the model is given heuristically. The data for the worked example came from a number of morning home to work trips carried out within the same time period in the City of Edmonton. From the empirical results it is concluded that the trip-makers sampled considered, with varying degrees of importance, the following factors in their choice between the two modes studied:

price of transit home-work trip,

time of transit home-work trip,

price of auto home-work trip,

time of atuo home-work trip,

daily parking costs.

The recommendations of the thesis apply to the calibration of the binary choice model, for a population, using a representative sample and are mainly concerned with the seemingly ever present problem of data collection.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the assistance given by:

Associate Professor J.J. Bakker of the Department of Civil Engineering at the University of Alberta, Edmonton, who suggested the research topic and who guided the course of this study.

Associate Professor A. Buse of the Department of Economics, University of Alberta, who gave much valuable advice concerning the methods used in this thesis.

The National Research Council, and the City of Edmonton, Traffic Engineering Department (with the approval of the City Commissioners), for financial assistance.

Mrs. Vivian Spak, for the excellent job done in typing this thesis.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Within an urban passenger transport system, containing a number of alternative modes of transportation, tripmakers use those modes which they, the tripmakers, consider to be the most convenient and efficient. The most predominantly used mode in North American cities today is the private automobile mode. As people demand more and more trips, planners are finding it increasingly more expensive to go on catering to the automobile. Since any city has only limited resources at its disposal, alternative cheaper forms of transport are going to have to be used. However, if these alternatives are going to be used they must be fiercely competitive to the private car. Thus it is in the interest of transportation system designers to identify and measure the principal factors which influence peoples' choice between modes, so that cheaper alternatives to the car mode can be designed to be efficient from the tripmakers' point of view.

STATEMENT OF THE PROBLEM

The aim of this thesis is to investigate the relationship between the mode-split of compulsory trips and the principal factors influencing it, e.g., characteristics of the tripmaker and of the transportation system. Work trips, which are compulsory, form the majority of trips during the morning and evening peak hours and it is with peak hour traffic in mind that this thesis was written. The data of the compulsory trips analyzed in this investigation came from a questionnaire survey especially conducted for this study. The

analytical technique developed gives the probability that a trip will be made by one mode over any other.

LIMITATIONS

While the principles outlined in this thesis can be applied to a system of many modes, it is concerned only with the bus transit and private automobile modes. Furthermore, only transportation system variables are considered and of these only trip price by auto and by bus, trip time by auto and by bus, and price of parking. The data used in the analysis were from home to work trips performed in early April 1970 by a number of people working in the Central Business District of the City of Edmonton. The sampling was biased and so the results of this thesis apply only to the tripmakers interviewed in the questionnaire survey. The work trips analyzed were part of the general morning home to work movement for the City of Edmonton which is one of the major movements borne by that city's passenger transport system.

The thesis goes on to introduce a passenger travel demand model, then criticizes its application to compulsory trips and then a binary choice model is introduced and applied to the data of the thesis.

CHAPTER II

REVIEW OF PREVIOUS WORK

What is referred to in this thesis, and by Wohl (Wohl, 1967), as the Kraft Demand Model (K.D.M.) was developed for a study, performed by the Systems Analysis and Research Corp. (SARC, 1963), on the intercity passenger traffic in the Washington-Boston Corridor. The main terms of reference of this study were:

1. Identification and Measurement of the Principal Factors Influencing Intercity Passenger Demand,
2. Identification and Measurement of the Principal Factors Influencing the Division of Intercity Passenger Demand by Mode (of transportation),
3. Projection of Intercity Passenger Demand in the Corridor through 1980.

In carrying out these objectives the researchers made use of the theory of demand from economics. Since they were only interested in intercity traffic, they aggregated individual tripmakers into their respective domicile cities and looked upon each city as a consumer of the commodity intercity passenger trips. As far as the individual tripmaker is concerned, transportation is a derived demand commodity, i.e., a commodity not wanted for its own sake. However, in developing the K.D.M., with each city as a consumer, transportation was considered to be a direct demand commodity, i.e., a commodity wanted for its own sake. The principal factors influencing a city's demand for intercity passenger trips were isolated and can be classified into two main

categories, viz., environmental and transportation. The environmental forces encompassed various socioeconomic attributes of the origin city (e.g. population) and destination city (e.g. recreational activity). The transportation forces consisted of the various properties (e.g. times, costs to user) of the transportation system serving the cities. Overall intercity passenger demand was divided up into demand by trip purpose (business and personal) and then subdivided into demand by the four main modes of transportation in the system (air, rail, bus, car). These separate demands were expressed as functions of the principal factors affecting them. These functions, known collectively as the K.D.M., are a group of direct demand models for intercity passenger trips, by trip purpose and by mode, of which the consumer is a city. The K.D.M. is expressed mathematically as follows:

$$N(i,j,i|P_o,M_o) = f[\underline{S}(i|P_o),\underline{A}(j|P_o),\underline{T}(i,j,i|P_o,M_o),$$
$$\underline{C}(i,j,i|P_o,M_o),\underline{T}(i,j,i|P_o,M_\alpha),$$
$$\underline{C}(i,j,i|P_o,M_\alpha)]$$

where

$N(i,j,i|P_o,M_o)$ = the number of round trips between origin city i and destination city j for purpose P_o by mode M_o,

$\underline{S}(i|P_o)$ = vector of socioeconomic characteristics appropriate to purpose P_o describing the travellers residing in origin city i,

$\underline{A}(j|P_o)$ = vector of socioeconomic and land-use characteristics describing the level of activity appropriate to

purpose P_o in destination city j,

$\underline{T}(i,j,i|P_o,M_o)$ = vector of travel time components for the round trip from origin i to destination j for purpose P_o by mode M_o,

$\underline{C}(i,j,i|P_o,M_o)$ = vector of travel cost components for the round trip between origin i and destination j for purpose P_o by mode M_o,

$\underline{T}(i,j,i|P_o,M_\alpha)$ = vector of travel time components for the round trip from origin i to destination j for purpose P_o by each of the alternative modes $(\alpha = 1,......,n)$, and

$\underline{C}(i,j,i|P_o,M_\alpha)$ = vector of travel cost components for the round trip between origin i and destination j for purpose P_o by each of the alternative modes $(\alpha = 1,......,n)$.

The model was calibrated for the cities in the Washington-Boston Corridor by making observations on the above variables for each city pair and by fitting this data to an arbitrary, logarithmic function by means of regression analysis. Thus objectives 1 and 2 of the study were completed. Objective 3, the projection of intercity passenger demand in the corridor through 1980, was accomplished by putting into the K.D.M. the expected future values of the independent variables assuming no structural changes in the model with time.

Previous transportation study methods treated separately trip generation, trip attraction, routing, and distribution of trips between modes. The significance of the K.D.M. is that it combines all these facets in one operation. An added advantage is that it can determine

both relative and absolute changes in demand due to changes in transportation system variables, whereas before only relative changes in demand could be determined.

Based on the relationships derived from the Washington-Boston Corridor data, it was estimated that Corridor travel would increase 195% in the period 1960 to 1980. It was also found that, of all the independent variables considered, auto cost had the least affect on demand for intercity trips.

The economic theory of demand on which the K.D.M. is based has been well tried and tested, inherent in this theory is the element of choice, both choice in the quantity demanded and choice between substitutes. Furthermore, in producing a demand function for a commodity, if there exists the element of choice and if the independent variables chosen are the prime affecters of demand then the demand function is valid. SARC (SARC, 1963) implicitly states that this element of choice held for the intercity trips considered in the Washington-Boston Corridor Study. For urban or intracity trips SARC (SARC, 1963) explicitly states that there is an independence between the demand for intracity trips and transportation system variables, and that there is little opportunity to avoid five or six round trips per week between home and work. Thus it would seem that the K.D.M., modified for urban use by splitting the city up into zones and considering interzonal traffic, is not suitable for the urban situation, particularly not for work trips. However, Domencich (Domencich, 1968) proposes the use of the K.D.M. for city conditions including a submodel for work

trips. He has one model for the morning and evening peak hours and one for non peak hour traffic. One model to describe traffic at any time of the day would involve some form of Spectral Analysis as outlined by Granger (Granger, 1964) and, as yet, has not been done. Considering those people who travel to and from work in the morning and evening peak hours respectively, then changes (assumed not to be drastic) in the transportation system would not be expected to affect their overall demand for work trips. Nor would any changes in the system be expected to draw more people into the labour pool and thereby alter the overall demand. Domencich (Domencich, 1968) writes, "While it is possible that changes in the transportation system will not affect trip generation, there is not good reason for making this assumption a priori. It seems better to avoid making this assumption altogether; we have everything to gain and nothing to lose by letting the question be settled through the results of empirical estimation." For the work trip submodel, at least, the transportation system variables could be assumed to affect choice between modes, and the origin and destination zones' variables could be considered as the elements instrumental in trip generation.

The bugbear of urban transportation is the peak hour work trip traffic. It is thus the priority of city transportation planners to determine how people choose between the various modes, so that a more efficient system can be designed to accomodate peak hour traffic. The mode-split for peak hour traffic could be determined from an appropriate intracity K.D.M. . In this model the reason for aggregating tripmakers into zones is threefold:

1. to be able to determine changes in demand for transportation due to changes (e.g. a change in population) in zonal properties,
2. to make an indivisible commodity divisible (a trip being an indivisible commodity and available only in whole units),
3. to make data manipulation more manageable (with a loss in accuracy).

The K.D.M. is usually calibrated using cross-section data, i.e., data collected in one time period. Thus for each zone a comprehensive survey of the tripmakers is required to determine the dependent and independent variables for that zone. Aggregation of data produces error which is unnecessary if all that is wanted is a measure of the mode-split. As will be shown in the next chapter a measure of the mode-split may be obtained from a binary choice model which is akin to a K.D.M. in which zone sizes have been reduced to one tripmaker. Thus all the data collected from the large sample survey can be used without any modification. This data would consist of tripmaker variables (e.g. income) as well as transportation system variables (of which the various costs are recommended to be user perceived costs) but no zone variables. The resulting model would give the probability of a tripmaker choosing, from between two modes, one of the modes by which to carry out his trip. The calibration of this binary choice model for a group of people is the investigation being reported in this thesis.

CHAPTER III

THEORY

The theory behind all economic demand models is, of course, the economic theory of demand. This theory concerns the relationship between the quantity purchased of, or demand for, a commodity, service or good by a consumer or market and the factors which influence this demand. The factors which influence demand for a commodity are things such as the price of the commodity, the prices of related commodities, income and taste of the consumer, and certain socio-demographic factors. A demand function is obtained when demand is expressed as a function of the variables influencing it. Regarding the demand function, the elasticity (e) of demand (D) with respect to the independent variable X is precisely defined as

$$e = \frac{\partial D/D}{\partial X/X} = \frac{X}{D}\frac{\partial D}{\partial X}$$

which is a ratio of proportional changes. It is expected, a priori, that elasticities with respect to variables directly related to the subject commodity or to complementary services, known as direct-elasticities (d), lie in the range

$$-\infty \leq d \leq 0.$$

It is also expected, a priori, that elasticities with respect to variables directly related to competing goods or substitutes, called cross-elasticities (c), fall in the interval

$$0 \leq c \leq \infty.$$

An example of a complement being parking and of a substitute being bus transport where car travel is the subject commodity. Each individual consumer has his own peculiar demand functions, which need not necessarily be constant with time. Looking upon the urban passenger transport system as a form of market in which buyers and sellers exchange trips, then the theory of consumer demand is appropriate for interpreting mathematically the urban transport system.

The demand function for a commodity is obtained by fitting a hyperplane to a collection of market statistics on transactions (i.e. quantity exchanged and the related values of variables influencing demand). However it should be realised that the above procedure may not result in a demand function. Not only does demand vary with the factors influencing it but so does supply and so the relationship derived by the above method may be a demand function, a supply function, or something between the supply and demand functions. The best possible situation for estimating the relationship between quantity exhanged and the factors influencing it which can be identified as a demand function is when the demand function is stable and the supply function is subject to great variability. What is being discussed in this paragraph is the Identification Problem which is discussed further in Klein (Klein, 1962).

The model reported in this thesis can be interpreted as the limiting case of the K.D.M. given in CHAPTER II, in which the large zonal aggregations have been reduced to zones of one work trip maker. Furthermore, only two modes, car and bus, were considered and the socioeconomic characteristics of the tripmakers have been omitted. The

form of the model is:

$$N(i|C) = f[P(i|C), P(i|B), T(i|C), T(i|B), P(i|P)]$$

where

$N(i|C)$ is the number of home to work trips, per morning by car, demanded by person i,

$P(i|C)$ is the price of the home to work trip by car for person i,

$T(i|C)$ is the journey time of the home to work trip by car for person i,

$P(i|B)$ and $T(i|B)$ are the same as $P(i|C)$ and $T(i|C)$ respectively except that they refer to the bus mode, and

$P(i|P)$ is the price of parking per workday for person i.

It is noted here that $N(i|C)$ can take only the values 1 or 0 depending on whether the mode is by car or by bus and in this way, it is posed, the model forces tripmakers' choice to be primarily between modes rather than the quantity of trips demanded.

A re-interpretation of the model is therefore needed and it is that it is not a demand function but a probability function. This type of model has been previously applied to urban travel by Warner (Warner, 1962) using data from the Chicago Area Transportation Study conducted around 1956. Although Warner (Warner, 1962) offers various functional forms for the probability or binary choice model, a linear form was chosen for the model of this thesis. The data used in this study came from a survey as described in the next chapter and the model was calibrated using regression techniques described in

CHAPTER VI. The dependent variable of the model gives the probability that a work trip will be conducted by car driver mode where the choice is between car driver mode and bus transit mode. The concept of elasticity of probability of choice is very similar to the concept of elasticity of demand given at the beginning of this chapter. A two-dimensional illustration of a binary choice curve is given in FIGURE 1, although the function of this thesis has six-dimensions and is of the form:

$$PR = A + B.PC + C.PB + D.TC + E.TB + F.PP$$

where

PR is the probability that car driver mode will be chosen for the work trip over bus transit mode,

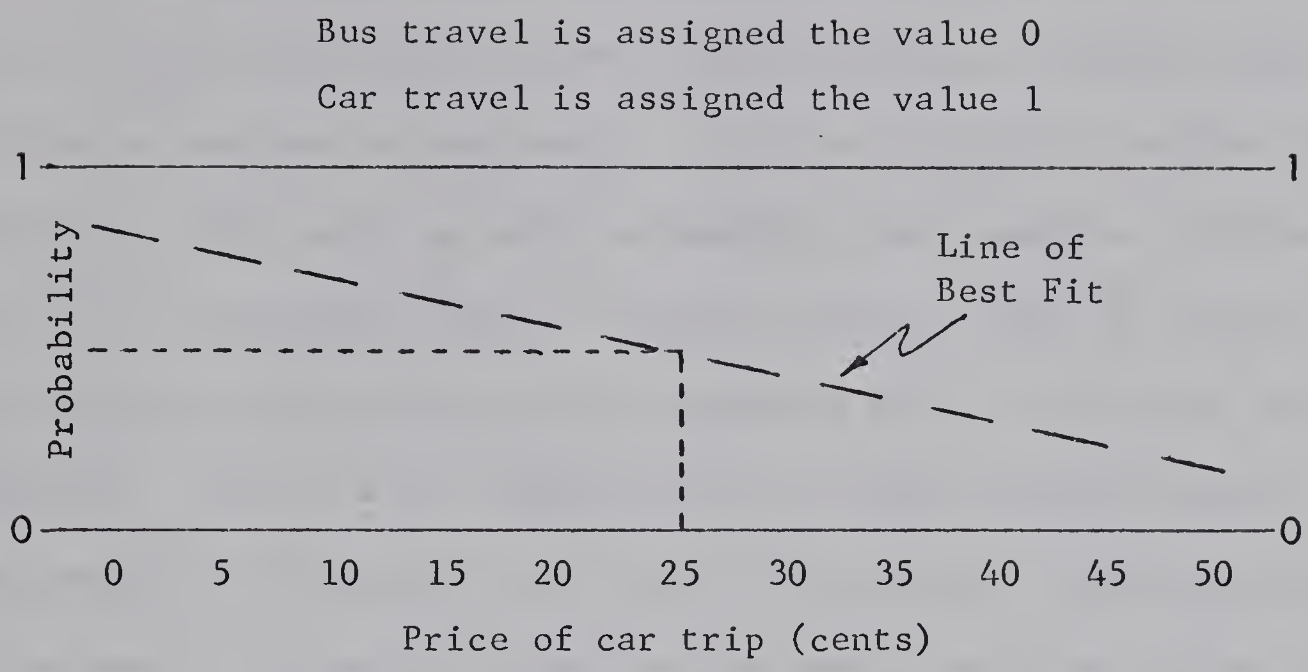
PC is the price of the auto home to work trip,

TC is the time of the home to work trip,

PB and TB are the same as PC and TC respectively except that they refer to the bus transit mode, and

PP are the daily parking costs.

The estimation of the parameters A,B,C,D,E and F, for a group of work trips, is described later in the thesis. From the car mode model a bus mode model may be obtained by reversing the signs on the regression coefficients and whose regression intercept is one minus the car model intercept.



It is expected, a priori, that most of the observations on bus travel will correspond to high car trip prices and that most of the observations on car travel will correspond to low car trip prices. Thus it is expected, a priori, that the line of best fit will have a slope in the direction as shown above. A car trip costing 25¢ in the above illustration has a probability of 50% that it will be carried out by car.

FIGURE 1:

ILLUSTRATION OF A BINARY CHOICE CURVE WHERE ONLY ONE INDEPENDENT VARIABLE AFFECTS THE CHOICE BETWEEN THE TWO MODES.

CHAPTER IV

PROCEDURE FOR DATA COLLECTION

The available data on the travel habits of the people living in the City of Edmonton were both incomplete, for the model of this thesis, and not contemporary. So it was decided to collect new information which would be both contemporary and complete. Furthermore, it is recommended that car running costs be what the drivers think they are and the only way to determine these is from the drivers themselves. Thus it became apparent that a questionnaire type of survey would be the easiest way to get the necessary quantities needed for the model. An initial survey on the people employed in the Royal Bank Building, City of Edmonton proved the feasibility of this method. This survey was carried out in the autumn of 1969 and the questionnaire used, entitled ROYAL BANK BUILDING SURVEY, is shown in APPENDIX A. The data from this survey could not be used in the model because the information given in each completed questionnaire referred only to one mode. However, the survey was useful in that a practical survey method was established and it gave what to expect in the way of demand for each of the modes listed on the survey form. The relevant facts from this survey are given in TABLE I.

The survey procedure used was as follows. Firstly, permission from the owner of the property was obtained, and then approximate weekly occupancy figures for the building were ascertained from the City of Edmonton, Planning Department. These figures were needed so as to know how many questionnaires to print. Each company registered in

TABLE I

ROYAL BANK BUILDING SURVEY INFORMATION

Mode	Male		Female		Total	
	No.	%	No.	%	No.	%
Car driver	183	74	12	6	195	42
Car passenger	15	6	54	25	69	15
Bus patron	36	15	134	61	170	37
Walker	10	4	16	7	26	5
Taxi patron	2	1	2	1	4	1
Total	246	100	218	100	464	100

Number of questionnaires given out = 611

Return = 464/611 = 76%

the building was then visited to explain to them the purpose of the survey and to get their approval to conduct it in their premises. Having obtained permission, a survey form for each employee was left to be filled in, and these were collected a few days after the first visit.

The data used to calibrate the model were obtained from a survey similar to the one above. It was carried out, within five days, in early April 1970 on three downtown Edmonton office towers, viz., the Centennial Building, Chancery Hall, and the Canadian National Tower. Only the three most popular modes of TABLE I, i.e., car driver, car passenger, and bus patron were thought to be pertinent. Thus the questionnaire used (shown in APPENDIX A and titled DOWNTOWN EDMONTON OFFICE BUILDING SURVEY) relates solely to these modes. A breakdown of the returned questionnaires by mode and other facts, appropriate to this stage of the thesis, concerning the survey are given in TABLE II. From this table it can be seen that, of the completed forms, 148/1233 = 12% were from car passengers. Originally, it was intended to construct models for the modes: car driver, car passenger, and bus patron, on the premiss that car passngers pay nothing for their trips. However, from the survey, this was found not to be so and, as there was no way of perceiving what a car driver or bus patron would pay for a car passenger trip and since car passengers formed only 12% of the sample, only the two other modes, i.e., car driver and bus patron, were considered. Hence the data referred to in the next chapter applies only to the 615 car drivers and 470 bus patrons who completed their questionnaires and none from the 148 car passengers who completed theirs.

TABLE II

DOWNTOWN EDMONTON OFFICE BUILDING SURVEY INFORMATION

	Centennial Building		Chancery Hall		C.N. Tower		Total	
	No.	%	No.	%	No.	%	No.	%
Car driver	128	45	79	45	408	53	615	50
male	116	41	71	41	372	48	559	45
female	12	4	8	4	36	5	56	5
Bus patron	129	45	70	40	271	35	470	38
male	62	22	23	13	205	27	290	24
female	67	23	47	27	66	8	180	14
Car passenger	29	10	25	15	94	12	148	12
Total	286	100	174	100	773	100	1233	100
Spoiled forms	248		121		264		633	
P	950		600		1100		2650	
Q	693		423		1330		2446	

Where P are the City of Edmonton, Planning Dept. figures and Q are the number of forms given out. Return = (615 + 470 + 148)/2446 = 52%.

CHAPTER V

OBSERVED DATA

The observed data from the 1085 (= 615 + 470) tripmakers who completed their survey forms were too bulky to be published in this thesis and so were lodged with the Civil Engineering Department, University of Alberta. However, the gist of the data can be seen from the means and standard deviations given in TABLE III.

The first things that can be said of the data are that they are complete and contemporary. That they are contemporary is because the information was collected in one time interval. That they are complete assures the correct relationship, for the beginning of April, between the variable values of each trip as these values were all taken at the same time. Furthermore, because the sample from which the data were drawn was biased, then, assuming the replies were honest appraisals, these data are only representative of the 1085 tripmakers surveyed. The sample is biased in the sense that it may not be indicative of the total home to work movement of the City of Edmonton. However it may very well be representative of the work trips to prestige office buildings in downtown Edmonton.

Regarding the separate means for variables (except parking costs) pertaining to car drivers and to bus patrons in TABLE III. It can be seen that the means of costs and times of the car and bus modes for car drivers are all greater than the corresponding means for bus patrons. The reason for this could be that:

TABLE III

MEANS AND STANDARD DEVIATIONS OF OBSERVED DATA

Variable	Car driver Mean	Bus patron Mean	Overall Mean	Standard Deviation
Demand for auto home-work trips	1.00000	0.00000	0.56682	0.49574
Price of auto home-work trip (cents	42.03740	36.69787	39.72441	31.26881
Price of transit home-work trip (cents)	23.94634	22.12979	23.15944	4.28757
Auto home-work journey time (mins)	32.82114	30.04468	31.61842	12.35649
Transit home-work journey time (mins)	50.88618	42.18298	47.11612	14.86761
Daily parking cost (cents)	50.22764	97.47872	70.69585	61.03905

a) on the whole, the car drivers lived further away than the bus patrons from their place of work,

or b) that the car drivers overestimated transit system variables while the bus patrons underestimated car driver mode variables,

or c) there could have been an intermixture of a) and b).

The separate means for daily parking costs given in TABLE III shows that the bus patrons paid, metaphorically speaking, nearly twice as much for parking as the car drivers. This could be an indication that, in aggregate, the car drivers enjoyed subsidized parking facilities whereas the bus patrons did not.

A histogram of the number of people who started their work trips at a given time is shown in FIGURE 2. This shows that all the people in the sample left home between 6:00 a.m. and 9:15 a.m. and that 89% of them left between 7:00 a.m. and 8:15 a.m. The average times of starting work by building were: Chancery Hall 8:28 a.m., Centennial Building 8:12 a.m., and the C.N. Tower 8:13 a.m., 93% of the people in the sample started work at or between 7:30 a.m. and 8:30 a.m.. The morning peak hour for the City of Edmonton contains the time interval from 7:30 a.m. to 8:30 a.m., so it is concluded that the majority of the trips sampled, terminated during the peak hour. TABLE IV shows the observed values for the variables considered and the number of observations on those values. The observed values given as ranges are, in the main, intervals containing a value on which most of the observations, ascribed to the range, were made.

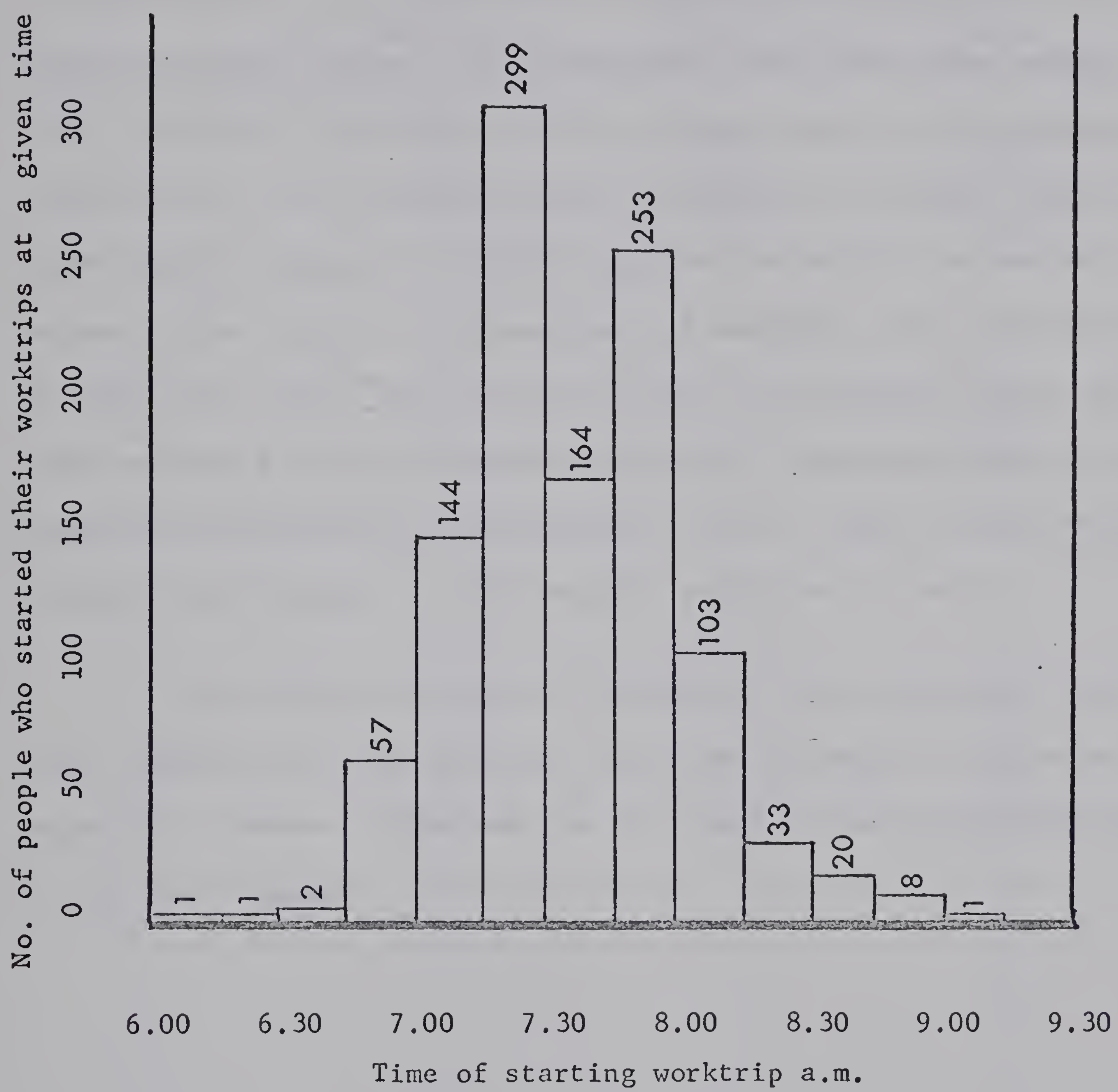


FIGURE 2:

HISTOGRAM OF PEOPLE LEAVING HOME FOR WORK
VERSUS TIME OF LEAVING

In regression analysis it is desirable to have the input data, for each variable, well distributed over the range from the smallest value to the largest. From the overall means and standard deviations of TABLE III it can be seen that the data for all variables are, to a greater or lesser extent, well distributed about their means except those for transit trip price. This is because most of the tripmakers reside in the City of Edmonton where the bus fare is fixed. TABLE IV shows how the data are distributed over the range from the smallest values to the largest. It shows that all variables have good distributions in their data except the transit trip price, as explained above, and the data associated with the dependent variable. The observed data for the dependent variable has no distribution along its range of observations because this variable can take only the values one or zero.

One has no quantitative conception of the accuracy of these data, indeed some of the journey times given in TABLE IV seem to be excessive. However, these were the data used in the analysis described in the next chapter.

TABLE IV

OBSERVED VALUES WITH CORRESPONDING NUMBERS OF OBSERVATIONS

Observed Value	Numbers of Observations		
	Car driver	Bus patron	Total
Dependent variable (dimensionless)			
0	0	470	470
1	615	0	615
	615	470	1085
Price of transit home-work trip (cents)			
0	1	2	3
20	237	289	526
25	323	116	489
30-50	54	13	67
	615	470	1085
Price of auto home-work trip (cents)			
0-2	5	3	8
3-7	10	10	20
8-12	33	27	60
13-17	51	32	83
18-22	58	49	107
23-27	97	99	196
28-32	56	48	104
33-37	32	10	42
38-42	27	34	61
43-47	11	7	18
48-52	117	93	210
53-57	7	3	10
58-62	17	11	28
63-67	8	5	13
68-72	5	3	8
73-77	20	12	32
78-82	13	1	14
83-87	2	0	2
88-92	3	2	5
93-97	2	0	2
98-102	21	15	36
103-410	20	6	26
	615	470	1085

TABLE IV (continued)

OBSERVED VALUES WITH CORRESPONDING NUMBERS OF OBSERVATIONS

Observed Value	Numbers of Observations Car driver	Bus patron	Total
Daily parking costs (cents)			
0-2	152	42	194
3-7	2	0	2
8-12	3	1	4
13-17	17	0	17
18-22	16	0	16
23-27	11	1	12
28-32	11	4	15
33-37	42	12	54
38-42	59	35	94
43-47	9	8	17
48-52	106	60	166
53-57	8	1	9
58-62	9	8	17
63-67	4	3	7
68-72	11	7	18
73-77	23	18	41
78-82	10	10	20
83-87	5	9	14
88-92	8	16	24
93-97	3	0	3
98-102	45	92	137
103-325	61	143	204
	615	470	1085
Time of transit home-work trip (mins)			
0-2	0	0	0
3-7	0	0	0
8-12	1	1	2
13-17	2	3	5
18-22	11	14	25
23-27	7	19	26
28-32	69	83	152
33-37	19	61	80
38-42	66	70	136
43-47	118	75	193
48-52	55	56	111
53-57	23	30	53
58-62	153	31	184
63-67	13	12	25
68-72	22	9	31
73-77	28	3	31
78-82	6	0	6

TABLE IV (continued)

OBSERVED VALUES WITH CORRESPONDING NUMBERS OF OBSERVATIONS

Observed Value	Numbers of Observations		
	Car driver	Bus patron	Total
Time of transit home-work trip (mins)			
83-87	4	1	5
88-92	13	2	15
93-97	0	0	0
98-102	5	0	5
	615	470	1085
Time of auto home-work trip (mins)			
0-2	0	0	0
3-7	3	0	3
8-12	9	10	19
13-17	54	48	102
18-22	72	82	154
23-27	38	40	78
28-32	182	162	344
33-37	63	23	86
38-42	71	31	102
43-47	54	46	100
48-52	23	7	30
53-57	8	3	11
58-62	31	13	44
63-67	1	0	1
68-72	4	0	4
73-77	2	2	4
78-82	0	2	2
83-87	0	0	0
88-92	0	1	1
	615	470	1085

CHAPTER VI

ANALYSIS AND RESULTS

The observed data from the 1085 car drivers and bus patrons, mentioned in the previous chapter, were used to build the model described in CHAPTER III. This model gives the probability that a work trip will be by car driver mode on the assumption that the independent variables used were the only ones affecting choice between the two modes considered.

Initially, an unconstrained multiple linear regression analysis was performed on the data to determine the signs and magnitudes of the unconstrained regression coefficients. The unconstrained regression analysis was performed by a program obtained from the University of Alberta, Department of Computing Science Program Library. The program was titled, "Multiple Linear Regression" (Number CS007), and was written by M. Easton of that department. The results of this analysis are given in TABLE V which shows that the sign of the coefficient for the price of the auto home-work trip is opposite to that expected a priori but that all other signs are as expected. (The overall averages and standard deviations of TABLE III and the t and F values of TABLE V were also part of the printout of program CS007). This necessitated the use of a regression analysis wherein the signs of the regression coefficients are constrained to those expected a priori. This type of regression analysis is defined as constrained regression analysis.

Constrained regression problems can be solved by means of a technique known as quadratic programming in which a quadratic function is optimized subject to various linear inequality constraints on the variables. The general quadratic programming problem is to choose w to optimize

$$p^t w + 0.5 w^t Q w$$

subject to the constraints

$$Rw \leq z,$$

$$w \geq 0.$$

In the above problem, lower case letters denote column vectors, upper case letters denote matrices, and t denotes transportation. Also, only w is variable and Q must be positive semi-definite, i.e. $v^t Q v \geq 0$ for all v. To facilitate solution, extra variables, known as slack variables, are introduced into vector w to convert the first inequality of the quadratic programming problem to an equality. These slack variables are governed by the second inequality and they have no affect on the function because their coefficients in the function are chosen to be zero.

Constrained regression analysis can be converted to a quadratic programming problem as shown in APPENDIX B. This problem is to choose x to minimize

$$(D^t D b_L - D^t y)^t x + 0.5 x^t D^t D x$$

subject to

$$x \leq b_U - b_L$$

$$x \geq 0.$$

In the above problem, as applied to this thesis, D is a 1085 × 6 matrix whose last five columns are observations on the independent variables and whose first column has elements all equal to one. Vector y has 1085 elements which are observations on the dependent variable, b_L and b_U are, respectively, the lower and upper bounds imposed on the regression coefficients. Vector x is defined as $x = b - b_L$, where b is the constrained regression coefficient vector whose first element is the intercept of regression. Thus once x has been determined, the regression coefficient vector, b, can be found from the relationship

$$b = x + b_L.$$

The particular quadratic programming problem for the data of this thesis was determined with the aid of the program listed in APPENDIX C. It is, on addition of slack variables, to choose vector x to minimize

$$542.5x_1^2 + 1386030.5x_2^2 + 300940.0x_3^2 + 625114.0x_4^2 + 1324116.5x_5^2 +$$
$$4730822.5x_6^2 + 43101.0x_1x_2 + 25128.0x_1x_3 + 34306.0x_1x_4 +$$
$$51121.0x_1x_5 + 76705.0x_1x_6 + 1027284.0x_2x_3 + 1393045.0x_2x_4 +$$
$$2098624.0x_2x_5 + 3178578.0x_2x_6 + 797810.0x_3x_4 + 1199788.0x_3x_5 +$$
$$1782071.0x_3x_6 + 1717546.0x_4x_5 + 2235668.0x_4x_6 + 3421588.0x_5x_6 -$$
$$154727.0x_1 - 7369537.0x_2 - 3621892.0x_3 - 4899126.0x_4 -$$
$$7269053.0x_5 - 14906781.0x_6 + 0.0x_7 + 0.0x_8 + 0.0x_9 + 0.0x_{10} +$$
$$0.0x_{11} + 0.0x_{12}$$

subject to

$$Ax = b_U - b_L$$
$$x \geq 0.$$

In the above problem the matrix A is

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix}$$

and b_U and b_L were chosen to be, respectively

$$(1,0,1,0,1,0)^t$$

and

$$(0,-1,0,-1,0,-1)^t.$$

The first element of vector x relates to the intercept of the constrained regression, the next five elements of vector x relate to the coefficients of the constrained regression, and the last six elements are slack variables. The problem was solved using the Convex Simplex Method (Zangwill, 1969), which is shown in algorithmic form in APPENDIX D. The FORTRAN program for this solution is listed in APPENDIX E where the initial basic feasible solution was taken to be

$$(1,1,1,1,1,1,0,0,0,0,0,0)^t.$$

In this program, step 3 case (b) and the first part of step 3 case (c) of the Convex Simplex Method (given fully in APPENDIX D) were deleted as they were not applicable in this case. (Since the elements of the tableau for each iteration were always greater than or equal to zero). The square of the coefficient of multiple correlation of the constrained regression analysis was determined using the program listed in APPENDIX F. This coefficient and the constrained regression results are shown in TABLE V. All programs were run in the system of the Department of Computing Science, University of Alberta.

The program listed in APPENDIX E did not reach the terminate condition of the Convex Simplex Method. With the round off error in the computer, this was expected, However, convergence was encountered on the 874th. iteration and the results of this iteration were used as a solution. Furthermore, the program made use of the fact that the inverse of a (square) matrix, each row and each column of which has elements all zero except for one element of value one, is equal to its

transpose.

In the previous chapter it was stated that one has no quantitative conception of the accuracy of the observed data. However, on the assumption that the replies were truthful, it is assumed that the data are free of measurement error and from now on the term error will be used for least squares residuals.

The square of the coefficient of multiple correlation of a regression analysis is taken as the register of compatibility between the observed dependent variable and that obtained by the imposed form of function operating on the observed independent variables. It ranges from greater than zero to one, the value one indicating perfect correlation. The relevant values for the constrained and unconstrained regressions of this thesis are 0.23130 and 0.23431 respectively which indicate poor fits.

The F value for the unconstrained regression analysis given in TABLE V indicates that this regression is significant at the 1% level. Since the constrained regression coefficients are very close to the unconstrained coefficients it is concluded that the constrained regression analysis is also significant at the 1% level.

The high error of the regression analyses is primarily due to the dependent variable taking only the values zero and one. Low values for coefficients of multiple correlation are the norm in economic cross-section analysis. More favourable figures could possibly be obtained for this thesis by using functional forms other than the linear one chosen. Better results for the coefficients of multiple correlation

may also be obtained by the inclusion of additional variables, e.g. income and car ownership, which were not considered in this treatment.

The figures given in TABLE V may not be indicative of the people who work in the Central Business District of the City of Edmonton because of the biased sampling. Thus the results apply only to the people in the sample. Regarding the unconstrained coefficient for the price of the auto home-work trip it is seen that its sign is opposite to that expected a priori. This looks like an example of the Giffen Paradox of economic literature. However it is more reasonable to suppose that with a fully specified model the signs of the regression coefficients would be as expected.

Again, since the constrained and unconstrained coefficients are so close it is assumed that the t values for the unconstrained regression given in TABLE V apply also to the constrained regression. At the 1% level of significance, all coefficients differ significantly from zero except the coefficient related to the price of the auto home-work trip. However, the coefficient of the price of the auto home-work trip differs significantly from zero at the 10% level of significance. This test is approximate because the standard errors of the coefficients vary with different values of the dependent variable whereas the t test assumes a constant relationship. However, the t test is usually applied in this way to linear probability functions.

The probability function derived for the tripmakers sampled is

$$P = 0.10400 + \underset{(0.000)}{0.00000PC} + \underset{(0.783)}{0.01915PB} - \underset{(-0.234)}{0.00419TC}$$

$$+ \underset{(0.641)}{0.00771TB} - \underset{(-0.372)}{0.00298PP}$$

where

P is the estimated probability of choice of car driver mode for the home-work trip,

PC is the price of the auto home-work trip (cents),

PB is the price of the transit home-work trip (cents),

TC is the auto home-work journey time (mins),

TB is the transit home-work journey time (mins),

PP is the daily parking cost (cents),

and where the figures in brackets are elasticities of probability of choice at the centroid of observations. Sensitivity of demand (Roth, 1967) with respect to an explanatory variable is the partial derivative of demand with respect to that variable. Thus the independent variable coefficients in the above function represent sensitivities of probability of choice of car driver mode for the home-work trip with respect to the respective independent variables. Of the three money variables considered the sensitivity associated with transit price is by far the largest. This suggests that for equal changes in these variables, transit price has the greatest affect on probability of choice of car driver mode over bus mode. Similarly, of the two time variables, transit time has the greater affect on the probability of choice. The sensitivity (and elasticity) associated with the price of the auto home-work trip is zero showing that changes in this variable do not cause changes in the dependent variable.

Considering now the elasticities given above, it can be seen that the elasticity related to the transit trip price is the largest of the five. Also associated with the transit mode is the next largest elasticity that concerning journey time by bus. The price of parking elasticity is the next largest and the two smallest elasticities relate to the car driver mode. Thus it seems that changes in transit system variables have the greatest affect, of the variables considered, on choice between the two modes.

A trip, whose explanatory variables are the same as those at the mean of observations, has a 43% probability that it will be performed by car driver mode. The results of this thesis cannot be compared with those given by Warner (Warner, 1962) as he has used different independent variables.

TABLE V

RESULTS OF REGRESSION ANALYSES

Variable	t Value	Regression Unconstrained	Coefficients Constrained
Dependent			
Probability of choice of auto mode			
Independent			
Price of auto home-work trip (cents)	1.93158	+0.00084	0.00000
Price of transit home-work trip (cents)	5.76555	+0.01860	+0.01915
Auto home-work journey time (mins)	- 3.25299	-0.00410	-0.00419
Transit home-work journey time (mins)	7.03533	+0.00752	+0.00771
Daily parking costs (cents)	-13.37453	-0.00300	-0.00298
Intercept of regression		+0.09009	+0.10400
Coefficient of multiple correlation		0.23431	0.23130
F value	66.03740		
Degrees of freedom	5 and 1079		

CHAPTER VII

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

This thesis investigated the relationship between mode-split for compulsory trips and the principal factors affecting it. It criticized a current passenger travel demand model with respect to compulsory trips and it went on to introduce a binary choice model which it was felt would be better for estimating mode-split for trips of necessity. The function derived did not give mode-split directly as it was actually a probability function which estimated the probability of a trip being made by a mode given a choice of two modes. The two modes considered were auto driver mode and bus transit mode. The factors affecting mode-split, or the choice between the two modes considered, were assumed to be trip price by auto and by bus, trip time by auto and by bus, and price of parking. The procedure for calibrating the model differed from the usual method of economic demand model building in that observations on the dependent variable were on two different commodities, viz., car travel and bus travel. The model of the thesis was constructed using information from 1085 work trips of which 615 were by car and 470 by bus. These data were obtained in a questionnaire survey and so the function derived pertains only to these trips.

Given a number of compulsory trips, performed on a number of modes and the explanatory variables related to each trip then the technique of this thesis can be used to determine the binary choice

function for each pair of modes.

On the assumption that the survey replies were honest then the following conclusions can be drawn from the results of this thesis:

1. Either a) the car drivers lived further away from their place of work than the bus patrons,

   or b) the car drivers overestimated transit system variables whereas bus patrons underestimated auto journey times and prices,

   or c) both a) and b) were operative.

This conclusion has been drawn from the results given in TABLE IV.

2. The tripmakers in the sample considered the money variables in the following decreasing order of influence in their decision to use either mode:

   Price of transit home-work trip,

   Daily parking costs,

   Price of auto home-work trip.

3. Changes in the time of the transit home-work trip were found to have a greater influence on choice between the two modes than similar changes in the time of the auto home-work trip.

The last two conclusions have been deduced from the results given in TABLE V.

RECOMMENDATIONS

1. Further research be conducted to determine the fullest possible extent of the principal factors influencing choice between modes.

2. Data collection methods other than questionnaire surveys be devised so that representative samples can be exploited to calibrate the model.
3. If perceived data are used to build the model then studies are needed into the correlation between perceived and arbitrarily measured quantities. The reason for this is to be able to determine the affect on mode-split of changes in transportation system variables which can be aribtrarily measured.
4. If arbitrarily measured data are used to build the model, then, since a large amount of data are needed, efficient, consistent, and accurate methods of data collection need be developed. Furthermore, the times and prices of the various modes should not be linear variations of distance, so that the problem of multicollinearity is avoided.
5. Since in practice it was found to be difficult to treat the car passenger mode as a separate mode, this mode and the associated car driver mode be treated as one mode, i.e., car pool mode.
6. Functional forms other than the linear one used in this thesis should be tried.
7. In addition to the independent variables used in this thesis, variables related to income, car ownership, age, sex, and trip length should be included in the model.

LIST OF REFERENCES

Domencich, T.A., Kraft, G., Valette, J-P., "Estimation of Urban Passenger Travel Behaviour: An Economic Demand Model", Highway Research Record Number 238, Washington, D.C., 1968.

Granger, C.W.J., "Spectral Analysis of Economic Time Series", Princeton University Press, Princeton, 1964.

Klein, L.R., "An Introduction to Econometrics", Prentice-Hall, Englewood Cliffs, 1962.

Roth, G., "Paying for Roads: The Economics of Traffic Congestion", Penguin Books, Harmondsworth, 1967.

SARC (Systems Analysis and Research Corp.), "Demand for Intercity Passenger Travel in the Washington-Boston Corridor", Report to the U.S. Dept. of Commerce, 1963.

Warner, S.L., "Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice", Northwestern University Press, Evanston, 1962.

Wohl, M., Martin, B.V., "Traffic Systems Analysis", McGraw-Hill, New York, 1967.

Zangwill, W.I., "Nonlinear Programming: A Unified Approach", Prenctice-Hall, Englewood Cliffs, 1969, APPENDIX D reproduced by permission of the publisher.

APPENDIX A

QUESTIONNAIRES USED IN SURVEYS

ROYAL BANK BUILDING SURVEY

The purpose of this survey is to study the effect of major office development on City growth and traffic patterns. The Royal Bank Building has been chosen as representative of this type of development. Your cooperation in completing the following questionnaire would be sincerely appreciated.

Question	Please put answers in this column
1) SEX: (M or F)	
2) ADDRESS: (nearest street & avenue)	
3) MODE OF TRAVEL TO WORK: a) car driver b) car passenger c) bus patron d) taxi e) walk f) other (please specify)	
4) IF YOU ARE A CAR PASSENGER DOES THE DRIVER WORK IN THE ROYAL BANK BUILDING:	
5) TIME OF LEAVING HOME: (to the nearest minute)	
6) TIME OF ARRIVAL AT WORK: (to the nearest minute)	
7) HOW MUCH, DO YOU ESTIMATE, ARE YOUR MONTHLY COSTS OF TRAVEL TO AND FROM WORK: (excluding parking costs if any)	
THE REMAINING TWO QUESTIONS ARE FOR CAR DRIVERS ONLY	
8) NEAREST STREET & AVENUE TO THE LOCATION OF YOUR PARKING STALL:	
9) OUT OF POCKET MONTHLY PARKING COSTS:	

Thank you for your cooperation

DOWNTOWN EDMONTON OFFICE BUILDING SURVEY

The purpose of this survey is to study the home to work travel habits of people employed in the City of Edmonton's Central Business District with a view to better understanding the City's transportation facilities. It is being conducted by the Civil Engineering Department of the University of Alberta as part of a continual research programme in urban transportation. Your co-operation in completing the following questionnaire would be sincerely appreciated.

	Please put answers in this column
1) ADDRESS: (nearest street & avenue)	
2) SEX:	M or F
3) PLEASE STATE WHETHER: a) Car driver b) Car passenger c) Bus patron	
4) TIME OF LEAVING HOME: (to the nearest minute)	A.M.
5) TIME OF STARTING WORK:	A.M.
6) COST: of your daily home to work journey excluding subsidies and parking costs if any:	¢ per trip
7) PARKING COST: for car drivers only, please exclude any subsidies:	¢ per day
8) CAR DRIVERS AND PASSENGERS: if you were to travel to work by bus what would be:	
a) Time you would have to leave home:	A.M.
b) Cost of the home to work journey:	¢ per trip
9) BUS PATRONS AND CAR PASSENGERS: if you were to drive a car to work each day what would be:	
a) Time you would have to leave home:	A.M.
b) Cost of the home to work journey:	¢ per trip
c) Daily parking cost:	¢ per day

Thank you for your co-operation

APPENDIX B

CONSTRAINED REGRESSION AS A QUADRATIC PROGRAMMING PROBLEM

The constrained least squares regression technique used in this thesis can be reduced to a quadratic programming problem by straightforward algebraic manipulation as shown below.

Matrices are represented by upper case letters, column vectors by lower case letters, superscript t denotes transposition.

The constrained regression problem is to choose b (a vector of regression coefficients) to minimize

$$e^t e$$

subject to

$$b_L \leq b \leq b_U, \tag{1}$$

where e, the vector of least squares errors, is defined by

$$e = y - Db. \tag{2}$$

In the inequality (1), b_L and b_U are, respectively, the lower and upper bounds imposed on the regression coefficients. For this particular problem there were 1085 observations and 5 independent variables. In equation (2), y is a 1085×1 vector of observations on the dependent variable, and D is a 1085×6 matrix whose last 5 columns are observations on the independent variables and whose first column has elements all equal to 1.0. The first element of the vector b is the intercept of the regression.

To convert the constrained regression problem into quadratic

programming format define

$$x = b - b_L. \tag{3}$$

The constraints then become

$$x \leq b_U - b_L \tag{4}$$

$$x \geq 0. \tag{5}$$

Substituting $e = y - Db$ in $e^t e$, the function to be minimized becomes

$$e^t e = y^t y - b^t D^t y - y^t Db + b^t D^t Db. \tag{6}$$

All the terms in equation (6) are scalars and so are equal to their transposes, therefore

$$e^t e = y^t y - 2b^t D^t y + b^t D^t Db. \tag{7}$$

Substituting $x = b - b_L$, and rearranging terms,

$$\begin{aligned} e^t e = {} & y^t y - 2x^t D^t y - 2b_L^t D^t y + x^t D^t Db_L \\ & + b_L^t D^t Dx + x^t D^t Dx + b_L^t D^t Db_L. \end{aligned} \tag{8}$$

Again all the terms in equation (8) are scalars and so equation (8) becomes

$$\begin{aligned} e^t e = {} & y^t y - 2x^t D^t y - 2b_L^t D^t y + 2x^t D^t Db_L \\ & + x^t D^t Dx + b_L^t D^t Db_L. \end{aligned} \tag{9}$$

In minimizing the right hand side of equation (9) a value of x is obtained which will be the same as that obtained by minimizing

the right hand side of (9) consisting only of those terms containing x. A new function to be minimized is now defined, $(e^te)^*$, which is derived from equation (9) by deleting those terms not containing x and by multiplying the resultant truncated form of e^te by 0.5.

$$(e^te)^* = -x^tD^ty + x^tD^tDb_L + 0.5x^tD^tDx. \tag{10}$$

The optimum value of x will be the same for both equations (9) and (10). Algebraic manipulation of equation (10) produces

$$(e^te)^* = (D^tDb_L - D^ty)^tx + 0.5x^tD^tDx. \tag{11}$$

Collecting results, the constrained least squares regression problem becomes the quadratic programming problem:

choose x to minimize

$$(D^tDb_L - D^ty)^tx + 0.5x^tD^tDx$$

subject to

$$x \leq b_U - b_L$$
$$x \geq 0.$$

Once x has been determined, the regression coefficient vector, b, can be found from the relationship

$$b = x + b_L.$$

There is a further requirement of the function to be minimized and it is that D^tD be positive semi-definite, i.e., $v^tD^tDv \geq 0$ for all v, which is the case for all D.

APPENDIX C

LISTING OF PROGRAM WHICH DERIVES THE QUADRATIC FUNCTION TO BE MINIMIZED

```
C THIS PROGRAM   REDUCES THE QUESTIONNAIRE DATA TO THE COEFF-
C ICIENTS OF X AND X**2 IN THE QUADRATIC PROGRAMMING PROBLEM
      DOUBLE PRECISION D(1085,6),BL(6),DTD(6,6),DTDBL(6),
     1DTY(6),COEFX(6),COEFXX(6,6)
      DO 10 I=1,1085
   10 READ(5,100) (D(I,J),J=1,6)
  100 FORMAT(6F13.0)
C DATA FROM QUESTIONNAIRES READ IN
      DATA BL/0.0,-1.0,0.0,-1.0,0.0,-1.0/
C BL READ IN
      DO 11 I=1,1085
   11 D(I,1)=1.0
C FIRST COLUMN OF D MODIFIED
      DO 12 N=1,6
      DO 12 M=1,6
      DTD(N,M)=0.0
      DO 13 I=1,1085
   13 DTD(N,M)=DTD(N,M)+(D(I,N)*D(I,M))
   12 CONTINUED
C DTRANSPOSE*D DETERMINED
      DO 14 I=1,6
      DTDBL(I)=0.0
      DO 15 J=1,6
   15 DTDBL(I)=DTDBL(I)+(DTD(I,J)*BL(J))
   14 CONTINUE
C DTRANSPOSE*D*BL DETERMINED
      DO 16 J=1,6
      DTY(J)=0.0
      DO 17 I=1,615
C THE FIRST 615 ELEMENTS OF Y ARE ONE THE REST ZERO
   17 DTY(J)=DTY(J)+D(I,J)
   16 CONTINUE
C DTRANSPOSE*Y DETERMINED
      DO 18 I=1,6
   18 COEFX(I)=DTDBL(I)-DTY(I)
C COEFFICIENT OF X DETERMINED
      DO 19 I=1,6
      DO 19 J=1,6
   19 COEFXX(I,J)= DTD(I,J)*0.5
C COEFFICIENT OF X**2 DETERMINED
      WRITE(6,101) (COEFX(I),I=1,6)
      WRITE(6,101) ((COEFXX(N,M),N=1,6),M=1,6)
  101 FORMAT(4X,D23.16)
      STOP
      END
```

APPENDIX D

CONVEX SIMPLEX METHOD

THE ALGORITHMIC PROCEDURE

The Simplex Method.

The simplex method for solving the Linear Programming problem is as follows: The L.P. problem is

$$\max q^t x$$

subject to

$$Ax = b$$
$$x \geq 0$$

where A is $m \times n$.

The simplex method operates on certain matrices T termed tableaux. Each tableau T has the property that

$$T = C^{-1}A$$

where C is an $m \times m$ basis matrix composed of m linearly independent columns of A. Thus if $A = (a^1,\ldots\ldots,a^m)$, letting a^i indicate a column of A,

$$C = [a^{N1},a^{B2},\ldots\ldots,a^{Bm}]$$

where Bj indicates a column number from A. The columns of A comprising C form the basis and are said to be basic. Observe that the Bjth column of T is a column of zeroes except for a one in the jth place.

Letting $b' = C^{-1}b$ it can be seen that any solution to

$$Tx = b'$$

is a solution to

$$Ax = b$$

and conversely.

A basic feasible solution is a point x such that

$$Tx = b'$$
$$x \geq 0$$

where in addition

$$x_{Bj} = b'_j \quad j = 1, \ldots\ldots, m$$

and all other x components are zero. The x_{Bj} correspond to the columns of A in the basis, and these x components are also said to be basic. Note that $b'_i \geq 0$ for all i.

Clearly, a basic feasible solution corresponds to a given tableau T.

Initialization Step.

An appropriate linear-simplex method phase I procedure has generated a basic f-asible solution x^1 with corresponding tableau T^1. Go to step 1 of iteration k with $k = 1$.

Iteration k.

The feasible point x^k and tableau $T^k = (t^k_{ij})$ are given.

Step 1.

Calculate the relative-cost vector: Let

$$c^k = (\nabla f(x^k) - (T^k)^t \nabla f(x^k)_B),$$

where

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n}\right),$$

$$c^k_{s1} = \min \{c^k_i \mid i \epsilon \nu\}$$

and

$$c^k_{s2} x^k_{s2} = \max \{c^k_i x^k_i \mid i \epsilon \nu\},$$

where ν is the set of the first n positive integers. If $c^k_{s1} = c^k_{s2} x^k_{s2} = 0$, terminate. Otherwise, go to step 2.

Step 2.

Determine the non basic variable to change:

If $c^k_{s2} x^k_{s2} \leq |c^k_{s1}|$, increase $x_s = x_{s1}$

If $c^k_{s2} x^k_{s2} \geq |c^k_{s1}|$, decrease $x_s = x_{s2}$

Adjust only basic variables.

Step 3.

Calculate x^{k+1}: There are three cases to consider.

Case (a) x_s is to be increased, and for some i, $t_{is}^k > 0$.

Increasing x_s will drive a basic variable to zero. Let y^k be the x value when that occurs. Specifically,

$$y_i^k = x_i^k \quad i \in N - s, \tag{1}$$

where $N - s$ is the set of indices of the non basic variables except s,

$$y_s^k = x_s^k + \Delta^k \tag{1a}$$

$$y_{Bi}^k = x_{Bi}^k - t_{is}^k \Delta^k \quad i \in \mu, \tag{1b}$$

where μ is the set of the first m positive integers, and

$$\Delta^k = x_{Br}^k / t_{rs}^k = \min \{x_{Bi}^k / t_{is}^k \mid t_{is}^k > 0\}.$$

Find

$$x^{k+1} \in M^1(x^k, d^k) \tag{2}$$

where $J = [0,1]$ and

$$d^k = y^k - x^k$$

i.e. $$f(x^{k+1}) = \min \{f(x^k + \tau d^k) \mid 0 \leq \tau \leq 1\}.$$

Case (b) x_s is to be increased, and $t_{is}^k \leq 0$ for all i.

In this case x_s may be increased indefinitely without driving a basic variable to zero. Define y^k as in equation (1) except let $\Delta^k = 1$. Determine x^{k+1} such that

$$x^{k+1} \in M^1(x^k, d^k)$$

where $J = [0,\alpha]$ for α very large and

$$d^k = y^k - x^k.$$

Case (c) x_s is decreased.

Determine y^k using equation (1) except define Δ^k as follows.

$$\Delta^k = \max \{\Delta_1^k, \Delta_2^k\}$$

where

$$\Delta_1^k = x_{Br}^k / t_{rs}^k = \max \{x_{Bi}^k / t_{is}^k \mid t_{is}^k < 0\}$$

and

$$\Delta_2^k = -x_s^k$$

should $t_{is}^k \geq 0$, $i \in \mu$, let $\Delta_1^k = -\infty$. Here y^k is the x corresponding to the point where, as x_s is decreased, either a basic variable becomes zero or x_s itself becomes zero, whichever occurs first. Calculate x^{k+1} using equation (2).

Step 4.

After x^{k+1} is calculated, determine tableau T^{k+1} as described in the next paragraph, and to to iteration k with $k + 1$ replacing k.

Tableau Selection.

In any of the above cases the tableau T^{k+1} is determined from x^{k+1} in the following manner. Order the components of x^{k+1} by magnitude:

$$x_{i1}^{k+1} \geq x_{i2}^{k+1} \geq \ldots\ldots \geq x_{in}^{k+1}$$

Determine the m largest in magnitude. Let T^{k+1} be any tableau such that these m are in the basis for that tableau. Thus $x_{i1}^{k+1}, \ldots\ldots, x_{im}^{k+1}$ forms the basis, and the corresponding columns of T^{k+1} are all zero except for a one in the appropriate position. Occasionally, due to linear dependencies, it may not be possible to place the largest m components in the basis. In such a case, determine a basis out of the largest $m + 1$ components of x^{k+1} or, if necessary, $m + 2$, $m + 3$, etc. The magnitude of the components of x^k in the basis should be as great as possible. In summary, place the largest number of the largest components of x^{k+1} in the basis.

APPENDIX E

LISTING OF PROGRAM WHICH SOLVES THE QUADRATIC PROGRAMMING PROBLEM BY THE CONVEX SIMPLEX METHOD

```
C THIS PROGRAM   IS A FORTRAN VERSION OF THE ALGROITHMIC
C PROCEDURE OF THE CONVEX SIMPLEX METHOD FOR THE SOLUTION
C OF THE PARTICULAR QUADRATIC PROGRAMMING PROBLEM DESCRIBED
C IN THIS THESIS
      DOUBLE PRECISION X(12),T(6,12),TMOD(6,12),SUMODT,
     1CBASIC(6,6),XBASIC(6),CINV(6,6),TT(12,6),DELFX(12),
     2DELFXB(6),TDFXB(12),C(12),CX(12),CK1,CXK2,ACK1,DEL,
     3Y(12),YBASIC(6),D(12),TAU,TAUINC,Z(12),FUNZ,FDATUM,
     4TNEW(6,12),BL(6),COEF(6)
      INTEGER I,J,K,L,M,N,KA,MU(6),K1,K2,LS,ITER
      DATA BL/0.0,-1.0,0.0,-1.0,0.0,-1.0/
      DATA X/6*1.0,6*0.0/
C BASIC FEASIBLE SOLUTION X1
      DO 10 I=1,6
      DO 10 J=1,12
   10 T(I,J)=0.0
      DO 11 N=1,6
      M=N+6
      T(N,N)=1.0
   11 T(N,M)=1.0
C TABLEAU T1 CORRESPONDING TO X1
C INITIALIZATION STEP COMPLETED
      ITER=0
C ITER IS ITERATION COUNTER
   12 ITER=ITER+1
      DO 13 I=1,6
      DO 13 J=1,12
   13 TMOD(I,J)=0.0
      DO 14 K=1,6
      L=K+6
      DO 15 M=1,6
      IF(X(K).GE.X(L)) TMOD(M,K)=T(M,K)
   15 IF(X(K).LT.X(L)) TMOD(M,L)=T(M,L)
   14 CONTINUE
      KA=1
      DO 16 I=1,12
      SUMODT=0.0
      DO 17 J=1,6
   17 SUMODT=SUMODT+TMOD(J,I)
      IF(SUMODT.EQ.0.0) GO TO 16
      DO 18 L=1,6
   18 IF(SUMODT.EQ.1.0) CBASIC(L,KA)=TMOD(L,I)
      XBASIC(KA)=X(I)
      MU(KA)=I
      KA=KA+1
   16 CONTINUE
      DO 19 I=1,6
      DO 19 J=1,6
   19 CINV(I,J)=CBASIC(J,I)
C INVERSE OF BASIS MATRIX DETERMINED
      DO 20 L=1,12
```

```
      DO 21 M=1,6
      TNEW(M,L)=0.0
      DO 22 N=1,6
   22 TNEW(M,L)=TNEW(M,L)+(CINV(M,N)*T(N,L))
   21 CONTINUE
   20 CONTINUE
      DO 23 I=1,6
      DO 23 J=1,12
   23 T(I,J)=TNEW(I,J)
C TABLEAU(K+1) FROM TABLEAU(K)
C STEP 4 COMPLETED
      DO 24 I=1,6
      DO 24 J=1,12
   24 TT(J,I)=T(I,J)
      DELFX(1)=(1085.0*X(1))+(43101.0*X(2))+(25128.0*X(3))+
     1(34306.0*X(4))+(51121.0*X(5))+(76705.0*X(6))-154727.0
      DELFX(2)=(2772061.0*X(2))+(43101.0*X(1))+(1027284.0*X
     1(3))+(1393045.0*X(4))+(2098624.0*X(5))+(3178578.0*X(6))
     2-7369537.0
      DELFX(3)=(601880.0*X(3))+(25128.0*X(1))+(1027284.0*X(2))
     1+(797810.0*X(4))+(1199788.0*X(5))+(1782071.0*X(6))
     2-3621892.0
      DELFX(4)=(1250228.0*X(4))+(34306.0*X(1))+(1393045.0*X
     1(2))+(797810.0*X(3))+(1717546.0*X(5))+(2235668.0*X(6))
     2-4899126.0
      DELFX(5)=(2648233.0*X(5))+(51121.0*X(1))+(2098624.0*X
     1(2))+(1199788.0*X(3))+(1717546.0*X(4))+(3421588.0*X(6))
     2-7269053.0
      DELFX(6)=(9461645.0*X(6))+(76705.0*X(1))+(3178578.0*X
     1(2))+(1782071.0*X(3))+(2235668.0*X(4))+(3421588.0*X(5))
     2-14906781.0
      DO 25 I=7,12
   25 DELFX(I)=0.0
      DO 26 I=1,6
   26 DELFXB(I)=DEFLX(MU(I))
C PARTIAL DERIVATIVES NOW IN SYSTEM
      DO 27 I=1,12
      TDFXB(I)=0.0
      DO 28 J=1,6
   28 TDFXB(I)=TDFXB(I)+(TT(I,J)*DELFXB(J))
   27 CONTINUE
      DO 29 K=1,12
   29 C(K)=DELFX(K)-TDFXB(K)
C RELATIVE COST VECTOR DETERMINED
      DO 30 L=1,12
   30 CX(L)=C(L)*X(L)
      CK1=C(1)
      K1=1
      DO 31 I=1,12
      IF(C(I).LT.CK1) GO TO 32
      GO TO 31
```

```
   32 CK1=C(I)
      K1=I
   31 CONTINUE
      CXK2=CX(1)
      K2=1
      DO 33 I=1,12
      IF(CX(I).GT.CXK2) GO TO 34
      GO TO 33
   34 CXK2=CX(I)
      K2=I
   33 CONTINUE
      IF(CK1.EQ.CXK2.AND.CK1.EQ.0.0) GO TO 51
C STEP 1 COMPLETED
      ACK1=DABS(CK1)
      IF(CXK2.LE.ACK1) go to 35
      IF(CXK2.GT.ACK1) GO TO 39
C STEP 2 COMPLETED
   35 LS=K1
      DO 36 I=1,6
      IF(T(I,LS).EQ.1.0) GO TO 37
      GO TO 36
   37 M=I
      GO TO 38
   36 CONTINUE
   38 DEL=XBASIC(M)
      GO TO 40
   39 LS=K2
      DEL=-(X(LS))
   40 DO 41 I=1,12
   41 Y(I)=X(I)
      Y(LS)=X(LS)+DEL
      DO 42 I=1,6
   42 YBASIC(I)=XBASIC(I)-(T(I,LS)*DEL)
      DO 43 J=1,6
   43 Y(MU(J))=YBASIC(J)
C VECTOR Y DETERMINED
      DO 44 I=1,12
   44 D(I)=Y(I)-X(I)
      TAU=0.0
      TAUINC=1.0
      DO 45 J=1,16
      TAUINC=TAUINC/10.0
      DO 46 I=1,20
      DO 47 L=1,12
   47 Z(L)=X(L)+(TAU*D(L))
      FUNZ=542.5*((Z(1))**2)+1386030.5*((Z(2))**2)+300940.0*
     1((Z(3))**2)+625114.0*((Z(4))**2)+1324116.5*((Z(5))**2)
     2+4730822.5*((Z(6))**2)+43101.0*((Z(L))*(Z(2)))+25128.0
     3*((Z(1))*(Z(3)))+34306.0*((Z(1))*(Z(4)))+51121.0*
     4((Z(1))*(Z(5)))+76705.0*((Z(1))*(Z(6)))+1027284.0*
     5((Z(2))*(Z(3)))+1393045.0*((Z(2))*(Z(4)))+2098624.0*
```

```
      6((Z(2))*(Z(5)))+3178578.0*((Z(2))*(Z(6)))+797810.0*
      7((Z(3))*(Z(4)))+1199788.0*((Z(3))*(Z(5)))+1782071.0*
      8((Z(3))*(Z(6)))+1717546.0*((Z(4))*(Z(5)))+2235668.0*
      9((Z(4))*(Z(6)))+3421588.0*((Z(5))*(Z(6)))-154727.0*
      1(Z(1))-7369537.0*(Z(2))-3621892.0*(Z(3))-4899126.0*
      2(Z(4))-7269053.0*(Z(5))-14906781.0*(Z(6))
C FUNZ IS FUNCTION(X+TAU*D) OR MAP1
       IF(TAU.EQ.0.0) FDATUM=FUNZ
       IF(FUNZ.GT.FDATUM) GO TO 48
       FDATUM=FUNZ
       TAU=TAU+TAUINC
   46 CONTINUE
       GO TO 45
   48 TAU=TAU-(TAUINC*2.0)
       DO 49 K=1,12
   49 Z(K)=X(K)+(TAU*D(K))
       FUNZ=542.5*((Z(1))**2)+1386030.5*((Z(2))**2)+300940.0*
      1((Z(3))**2)+625114.0*((Z(4))**2)+1324116.5*((Z(5))**2)
      2+4730822.5*((Z(6))**2)+43101.0*((Z(1))*(Z(2)))+25128.0
      3*((Z9L))*(Z(3)))+34306.0*((Z(1))*(Z(4)))+51121.0*
      4((Z(1))*(Z(5)))+76705.0*((Z(1))*(Z(6)))+1027284.0*
      5((Z(2))*(Z(3)))+1393045.0*((Z(2))*(Z(4)))+2098624.0*
      6((Z(2))+(Z(5)))+3178578.0*((Z(2))*(Z(6)))+797810.0*
      7((Z(3))*(Z(4)))+1199788.0*((Z(3))*(Z(5)))+1782071.0*
      8((Z(3))*(Z(6)))+1717546.0*((Z(4))*(Z(5)))+2235668.0*
      9((Z(4))*(Z(6)))+3421588.0*((Z(5))*(Z(6)))-154727.0*
      1(Z(1))-7369537.0*(Z(2))-3621892.0*(Z(3))-4899126.0*
      2(Z(4))-7269053.0*(Z(5))-14906781.0*(Z(6))
       FDATUM=FUNZ
   45 CONTINUE
C FUNZ MINIMIZED BETWEEN TAU = 0.0 AND TAU = 1.0
       IF(TAU.LE.0.0) GO TO 51
       IF(TAU.GE.1.0) TAU=1.0
       DO 50 I=1,12
   50 X(I)=X(I)+(TAU*D(I))
C STEP 3 COMPLETED
       GO TO 12
C X(K+1) DETERMINED GO ON TO NEXT ITERATION
   51 DO 52 I=1,6
   52 COEF(I)=X(I)+BL(I)
C CONSTRAINED REGRESSION COEFFICIENTS DETERMINED
       WRITE(6,100) ((COEF(I),I=1,6),ACK1,CXK2,TAU)
  100 FORMAT(///30X,D23.16)
       WRITE(6,101) ITER
  101 FORMAT(///30X,I6)
       STOP
       END
```

APPENDIX F

LISTING OF PROGRAM WHICH DETERMINES RSQUARED FOR THE CONSTRAINED REGRESSION ANALYSIS

```
C THIS PROGRAM   DETERMINES THE COEFFICIENT OF MULTIPLE
C CORRELATION (RSQUARED) FOR THE CONSTRAINED REGRESSION
C PROBLEM
      DOUBLE PRECISION D(1085,6),Y(1085),B(5),YBAR,SUMYSQ,
     1XBAR(6),SUMCSQ,RSQ
      DO 10 I=1,1085
   10 READ(5,100) (D(I,J),J=1,6)
  100 FORMAT(6F13.0)
C QUESTIONNAIRE DATA READ IN
      DO 11 I=1,1085
   11 Y(I)=D(I,1)
C Y IS THE DEPENDENT VARIABLE VECTOR
      B(1)=0.0
      B(2)=0.0191524680425652
      B(3)=-0.004193437786666l
      B(4)=0.00771419122028840З
      B(5)=-0.002984478749278
C CONSTRAINED REGRESSION COEFFICIENTS READ IN
      YBAR=0.0
      DO 12 I=1,1085
   12 YBAR=YBAR+Y(I)/1085.0
      SUMYSQ=0.0
      DO 13 I=1,1085
   13 SUMYSQ=SUMYSQ+((Y(I)-YBAR)**2)
      DO 14 J=1,6
      XBAR(J)=0.0
      DO 15 I=1,1085
   15 XBAR(J)=XBAR(J)+(D(I,J)/1085.0)
   14 CONTINUE
      SUMCSQ=0.0
      DO 16 J=2,6
      K=J-1
      DO 16 I=1,1085
   16 SUMCSQ=SUMCSQ+(B(K)*((Y(I)-YBAR)*(D(I,J)-XBAR(J))))
      RSQ=SUMCSQ/SUMYSQ
C RSQUARED DETERMINED
      WRITE(6,101) RSQ
  101 FORMAT(50X,D23.16)
      STOP
      END
```